The Pharmacogenomics Journal (2001) 1, 161–166

www.nature.com/tpj

Chipping away at the mystery of drug responses

JC Rockett

Reproductive Toxicology Division, National Health and Environmental Effects Research Laboratory, United States Environmental Protection Agency, Research Triangle Park, NC, USA

DNA arrays can be used to discriminate closely related genes and provide a facile overview of gene expression responses relevant to drug metabolism and toxicology.

The figures are startling—in the United States alone, more than 100 000 die every year from the side effects of properly prescribed medicines, and a further two million develop serious side effects.[1] The new shield to be hefted against such collateral damage is genomics. One of the most widely touted suitors for genomics has been pharmacology, where it has been welcomed as a tool to reprieve the increasingly costly process of bringing new drugs to market and remove some of the guesswork from drug prescription. Indeed, it is widely believed that pharmacogenomics, the hybrid offspring of the two disciplines, will in time lead to the holy grail of medicine—the treatment of patients on a case-by-case basis, utilizing their genotype as a means to predict which drugs will be effective and which will produce adverse outcomes. Adverse outcomes are frequently attributed to genetic variability. In some cases, drugs have no effect because patients carry out ultra-rapid clearance as a result of amplification of the genes that metabolize them. In contrast, toxic side-effects are usually considered a result of sub- or non-functional genes which are unable to metabolize certain drugs, a situation often linked to the possession of certain gene polymorphisms.

With many genome projects completed and more soon to follow, focus is beginning to shift from gene discovery to the more complex arena of gene expression and function. In an article recently published in *Physiological Genomics* by Gerhold *et al*,[2] the authors confirm what many in the field already 'knew', or at least strongly suspected—that DNA arrays are indeed useful tools for characterizing gene expression changes induced by drugs, and that these changes are distinct enough to differentiate between drug classes. The study is not terribly unique in that others have already used arrays to detect cytochrome P450 (CYP) genes,[3] confirmed gene changes detected by Affymetrix microarrays[4] and conducted proof-of-principle studies on the ability of arrays to discriminate between different drug and chemical responses.[5–9] Nevertheless, Gerhold's experiments were carried out using well-characterized, mechanistically distinct model compounds (3-methylcholanthrene, clofibrate, dexamethasone and phenobarbitol) in a classic exposure model (*in vivo* rat liver), and the reported findings are well supported by the literature. The genes with altered expression included all the usual suspects (CYPs, glutathione S-transferases, UDP-glucuronyl transferases etc), and thus provide further evidence for pharmacologists that arrays can indeed 'provide a facile overview of gene expression responses relevant to drug metabolism and toxicology'. And, whilst the primary aim of the Gerhold study was to investigate the feasibility of using array technology to speed up the drug candidate selection process, several other aspects of the work are worthy of further comment from a pharmacogenomic perspective.

At the heart of array technology are the genes that make up the array. The right genes must be selected to answer the question being asked. Though many researchers who have crafted purpose-specific arrays might argue otherwise, gene selection for such arrays is still a somewhat arbitrary process, being biased by the knowledge (or lack thereof) of the developers, and restricted by the incomplete pool of known genes and a lack of understanding of their functional profile. Until cellular signaling pathways leading to toxicity are thoroughly characterized and the key genes in drug response pathways identified, it will always be a shot in the dark, or at least the gloom, to develop a chip for all seasons. Thus, although the so-called 'Drug Safety Chip' may represent Merck's current best shot, more appropriate gene sets will undoubtedly be deposited in subsequent iterations of their chip (and similar ones developed by companies such as Phase-1 Toxicology[9]) as new genes are characterized and toxic mechanisms elucidated.

Gerhold *et al* chose to use steady state dose models with their selected drugs. This approach is reasonable in a proof-of-concept study, where the aim is simply to differentiate drug responses at the RNA expression level. However, applied pharmacogenomics requires some kind of toxicological or pharmacological endpoint to compare with the expression profiles, so that a prognostic or diagnostic relationship can be established between them. Ideally, expression patterns generated by 'no observed adverse effect level' (NOAEL), 'lowest observed adverse effect level' (LOAEL) and LD_{50} doses should be compared to deduce if different genes and pathways are activated at the non-toxic, toxic and lethal doses which cause, or are corre-

lated with, observed effects. Interanimal variation is an important consideration in the selection of such dose levels. Preliminary studies by the authors indicated that interanimal variation exceeded experimental variation, and this was used to support pooling of samples for further analysis. Although this approach may be applicable for mechanistic studies, there would appear to be less value from the pharmacogenomic standpoint. Individual variation, which exists even in inbred strains, lies at the very heart of pharmacogenomics, and should be embraced rather than smoothed away. Interestingly, the authors also pointed out that some genes were hypervariable among individual animals. Possible reasons for this were not offered, but are most likely caused by polymorphisms in the gene regulatory regions. One of the challenges for pharmacogenomics, therefore, is to identify and characterize the source of such differences and how they relate to drug metabolism.

The most common polymorphism is the single nucleotide polymorphism (SNP), and it has been estimated that there may be as many as 200–300 000 of them in the protein coding sequences of the human population. A large school of thought supports the contention that characterizing these SNPs is the key to deciphering the genetic basis of complex disease. However, a more realistic first step for pharmacogenomics is to properly decipher drug response at the gene level. For example, the CYP family of genes is of central importance in drug metabolism, and several members have gene sequences over 90% homologous. Furthermore, many of the CYP genes have multiple allelic variants that are responsible for a high proportion of observed variations in drug response. It is therefore a significant challenge to characterize which of these genes and alleles are regulated in response to specific drug exposures. RT-PCR is probably the best current method for discriminating the expression of such drug metabolizing genes, as it provides both quantitative expression data and can differentiate allelic variants.[10] Unfortunately, RT-PCR is a relatively low-throughput technology and researchers are seeking ways to replace it with higher throughput microarrays. However, there is concern that the widely used cDNA arrays are unable to discriminate such closely related genes. There is a move, therefore, towards arrays composed of short (50–80 base) oligonucleotides. These can, with careful selection, be constructed from unique regions of the chosen genes to theoretically provide good discriminatory power. However, since the overall efficacy of this length of probe is still under scrutiny, the Affymetrix platform of multiple, short oligonucleotides per gene remains the current gold standard. Gerhold's paper demonstrates the ability of Affymetrix chips to distinguish members of the CYP family with >90% sequence homology. This kind of resolving power will be hard to beat without a jump in technology, and may permit Affymetrix to develop a very firm grip on this area of the market.

Those working with arrays learn quickly that deciphering the data is the key to the treasures that lie hidden within, and that the key is hard to turn. Not surprisingly then, these industrially based researchers took the logical step of designing a custom database to hold and mine their data. This is all well and good for private companies with proprietary concerns, but raises the wider issue of just how much useful information remains locked up in the archives of such private entities, untapped by their curators because of lack of knowledge or interest. The development of a national or international gene expression database is currently a difficult prospect due to such factors as lack of cross-platform concordance, the disarrayed state of gene nomenclature and annotation, and a jealous desire to guard data from competitors. However, it can be argued that it would be in the best interests of medical research for gene expression laboratories to pool data in a central database(s), perhaps after they have extracted information that interests them. In this vein, the European Bioinformatics Institute has set up the Microarray Gene Expression Database group,[11,12] a cross-organization entity whose goal is 'to facilitate the establishing of gene expression data repositories, comparability of gene expression data from different sources and interoperability of different gene expression databases and data analysis software'. It should also be of no surprise if pharmaceutical companies soon start scrambling to form strategic alliances with one another in order to reduce experimental costs and mine the untapped wealth of one another's data.

Although achieving the goal of using gene expression profiles for rapid and early screening of new drugs is still a little way off, the data provided by Gerhold *et al* and others like it is clearly encouraging, and indicates that the time when this approach will be used routinely to aid in drug development is not that far away.

ACKNOWLEDGEMENTS

This document has been reviewed in accordance with the US Environmental Protection Agency policy and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use. Thanks to Drs Mitch Rosen (US EPA) and Ian Dix (AstraZeneca) for scientific review of this manuscript prior to submission.

DUALITY OF INTEREST

None declared.

Correspondence should be sent to
JC Rockett, Reproductive Toxicology Division (MD-72), National Health and Environmental Effects Research Laboratory, US EPA, Research Triangle Park, NC 27711, USA.
Tel: +1 919 541 2678
Fax: +1 919 541 4017
E-mail: rockett.john@epa.gov

1 Lazarou J *et al. JAMA* 1998; **279**: 1200–1205.
2 Gerhold D *et al. Physiol Genomics* 2001; **5**: 161–170. http://physiolgenomics.physiology.org/cgi/content/full/5/4/161
3 Nguyen L *et al. Drug Metab Dispos* 2000; **28**: 987–993.
4 Jelinsky S, Samson L. *Proc Natl Acad Sci USA* 1999; **96**: 1486–1491.
5 Giaever G *et al. Nat Genet* 1999; **21**: 278–283.
6 Scherf U *et al. Nat Genet* 2000; **24**: 236–244.

7 Waring J *et al*. *Toxicol Lett* 2001; **120**: 359–368.
8 Bartosiewicz M *et al*. *Environ Health Perspect* 2001; **109**: 71–74.
9 Burczynski M *et al*. *Toxicol Sci* 2000; **58**: 399–415.
10 Roberts R *et al*. *Hum Mutat* 2000; **16**: 77–85.
11 www.mged.org
12 Kellam P. *Genome Biol* 2001; **2**: REPORTS4011.

Jets: a modification to speed flexible oligonucleotide array construction

B Yoder[1] and SM Sell[2]

[1]*Department of Cell Biology, University of Alabama at Birmingham, AL, USA:* [2]*Human Genetics and Nutrition Sciences, University of Alabama at Birmingham, AL, USA*

We have come a long way since the days of Northern blots. With the completion, or near completion, of multiple genome sequencing projects, DNA microarray technology is emerging as an extremely powerful and fundamental tool that can be used to explore genetic and molecular pathways associated with disease and other cellular processes.[1,2] Translation of this information to the clinical arena will result in improvements in patient diagnostics and treatments.[3–5] The application of array technology to biological questions is quickly transforming scientific paradigm, from one that formerly centered on the analysis of a relatively few biological measurements to one that supports the simultaneous exploration of thousands of events. While array technology is a powerful tool for both researchers and clinicians, its utility is hindered by the expense of, and difficulties associated with, custom modification of the arrays. In a recent article published in *Nature Biotechnology*, Hughes *et al* describe the adaptation of ink-jet technology for the *in situ* synthesis of small quantities of unique 60-mer oligonucleotides directly onto glass slides.[6] This 'second generation' ink-jet oligonucleotide synthesizer is based on the approach described by Blanchard *et al*.[7] It will greatly facilitate array modification by eliminating the time necessary for large-scale cDNA or oligonucleotide synthesis, reorganization of microtiter dishes, re-spotting of the DNA elements, or the generation of new photolithographic masks. With this new application of ink-jet technology, small quantities of oligonucleotides, specified by the user, can be synthesized directly onto the array. This permits a rapid customization that can meet the widely varying needs of investigators.

The prototypic microarray contains thousands of DNA elements (cDNAs or oligonucleotides) that are spotted or synthesized onto a solid substrate.[8] These DNA elements correspond to the sequence of known genes (or predicted genes) identified through the efforts of the genome projects. Microarrays are typically constructed using two different approaches. The first approach was described by Pat Brown of Stanford University and involves mechanical spotting of cDNA fragments or oligonucleotides (60 to 70-mers) onto a solid substrate.[9] Deposition of the DNA is accomplished using robotics equipped with micropins or by ink-jet print heads; small volumes are transferred from microtiter wells to the surface of the slide.[10] While researchers using this 'spotting' approach are able to customize the array format in order to meet the laboratories' needs, this modification is cumbersome and requires continuous, large-scale production of new reagents (oligonucleotide synthesis or PCR amplification), the generation of new stock titer plates, and respotting.

The second approach, described by Fodor *et al*, and commercialized by Affymetrix, Inc, utilizes oligonucleotides (generally 25-mers) that are synthesized *in situ* onto a glass wafer via a photolithographic activation and masking process.[11,12] A stepwise series of nucleotide additions and light directed deprotection and chemical linking steps stipulate where a specific nucleotide is to be added on the growing oligonucleotide chain. Thus, in contrast to the rapid ink-jet mediated synthesis approach that requires a signal step to add each base to the growing oligonucleotide, the Affymetrix-based GeneChip™ photolithographic method requires four successive cycles (for each of the four possible nucleotides) to add each base to the oligonucleotide. The photolithographic process utilized by Affymetrix Inc does allow construction of very high-density oligonucleotide arrays containing spots as small as five microns. A major quandary with the Affymetrix-based technology is that each array requires a new set of unique photolithographic masks to define the specific areas on the array where an additional monomer is to be added. Thus, an array containing 60-mers would require 240 unique masks. While the Affymetrix-based technology is powerful and sophisticated, the format is not easily customized without incurring significant cost. This is due to the necessity to generate a large number of photolithographic masks and the extra time that it takes to synthesize the new oligonucleotides.

The approach utilized by Hughes *et al*[6] is a modification of an existing methodology described by Blanchard *et al*[7] and Southern *et al*.[13] The studies conducted by Southern *et al* in 1994 relied on small mechanical seals to define the site of oligonucleotide synthesis. While their data established the feasibility of miniaturizing the *in situ*